Question 1

(a) Half of all messages are spam: junk mail that you don’t want to read. A spam filter is a piece of software that tries to filter these spam messages before they reach your inbox. Suppose you have a spam filter that correctly detects 95% of spam messages, with only a 1% probability of a false positive — a message being marked as spam when it isn’t. If a message is flagged by the filter as spam, what is the probability that it is in fact not spam? [2 marks]

(b) The probability that a sports fan supports Sportball Team A is 0.4. Using a binomial distribution find the probability that in a randomly selected sample of 8 fans there are:

(i) exactly 3 who support Team A;

(ii) more than 5 who support Team A. [3 marks]

Question 2

A delivery driver travels $K_1$ kilometres on Tuesday, $K_2$ km on Wednesday, and $K_3$ km on Thursday. Here $K_1$, $K_2$, and $K_3$ are independent random variables each having a normal distribution with mean 90 and standard deviation 25.

(a) Calculate the probability that on Tuesday the driver travels more than 100 km.

(b) Calculate $P(80 \leq K_2 \leq 100)$.

(c) Calculate the probability that the driver travels less than 80 km on every one of the three days.

(d) Random variable $T = K_1 + K_2 + K_3$ records the total distance in kilometres travelled over the three days.

Calculate the expectation, variance, and standard deviation of $T$.

(e) Random variable $A = T/3$ records the average distance in kilometres travelled per day.

Calculate the expectation, variance, and standard deviation of $A$.

Show your working for each calculation. Look in Appendix A of the course textbook, Carlton and Devore, for Table A.3 on pages 601 and 602 showing the CDF $\Phi$ of the Standardized Normal distribution. [5 marks]
Solution 1

(a) We define events: $S$ is the event that a message is spam; $S^c$ is the event that a message is not spam; $D$ is the event that the spam filter flags a message as spam. We are given the following data.

$$P(S) = P(S^c) = \frac{1}{2}$$  
Half of all messages are spam; half are not.

$$P(D|S) = \frac{95}{100}$$  
The filter detects 95% of spam messages.

$$P(D|S^c) = \frac{1}{100}$$  
The filter marks 1% of non-spam messages as spam.

We can then reason using Bayes to find the probability we want.

$$P(S^c|D) = \frac{P(S^c \cap D)}{P(D)} = \frac{P(D|S^c)P(S^c)}{P(D|S)P(S) + P(D|S^c)P(S^c)}$$

$$= \frac{0.01 \cdot 0.5}{0.95 \cdot 0.5 + 0.01 \cdot 0.5}$$

$$= \frac{1}{96} = 0.010 \quad (3 \text{ d.p.})$$

For any message marked as spam there is a 1% probability that it is in fact not spam.

It is also possible to work this out using a tree diagram labelled with all the relevant probabilities.

(b) To treat this using a binomial distribution we regard “Supporting Team A” as success, giving $p = 0.4$ and $q = (1 - p) = 0.6$. Let $X$ be the random variable “number of Team A supporters in the random sample of 8 fans”. We then have $X \sim \text{Bin}(8, 0.4)$ and so

$$P(X = k) = \binom{n}{k} p^k q^{n-k} = \binom{8}{k} 0.4^k 0.6^{n-k}, \quad k = 0, 1, \ldots, 8.$$

We can then use this formula to calculate the required values.

(i)  
$$P(X = 3) = \binom{8}{3} (0.4)^3 (0.6)^5 = 0.279 \quad (3 \text{ d.p.})$$

(ii)  
$$P(X > 5) = P(X = 6) + P(X = 7) + P(X = 8)$$

$$= \binom{8}{6} (0.4)^6 (0.6)^2 + \binom{8}{7} (0.4)^7 (0.6)^1 + \binom{8}{8} (0.4)^8 (0.6)^0$$

$$= 0.050 \quad (3 \text{ d.p.})$$
Solution 2

(a) \( X \sim N(90, 25) \) where \( X \) is \( K_1, K_2 \) or \( K_3 \).

\[
P(K_1 > 100) = P\left(Z > \frac{100 - 90}{25}\right) = P(Z > 0.4) = 1 - \Phi(0.4) = 0.345 \quad (3 \text{ d.p.})
\]

(b) \[
P(80 \leq K_2 \leq 100) = \Phi(0.4) - \Phi(-0.4) = 0.311 \quad (3 \text{ d.p.})
\]

(c) \[
P(K_1 < 80 & K_2 < 80 & K_3 < 80) = P(K_1 < 80) \cdot P(K_2 < 80) \cdot P(K_3 < 80)
= \Phi(-0.4)^3 = 0.041 \quad (3 \text{ d.p.})
\]

(d) \[
E(T) = E(K_1) + E(K_2) + E(K_3) = 270
\]
\[
\text{Var}(T) = \text{Var}(K_1) + \text{Var}(K_2) + \text{Var}(K_3) = 3 \times 25^2 = 1875
\]
\[
\text{SD}(T) = \sqrt{\text{Var}(T)} = 43.3 \quad (3 \text{ s.f.})
\]

The first of these relies on linearity of expectations, the second depends on the \( K_i \) being independent, and the third follows from the definition of standard deviation.

(e) \[
E(A) = E(T)/3 = 90
\]
\[
\text{Var}(A) = \text{Var}(T)/3^2 = 208 \quad (3 \text{ s.f.})
\]
\[
\text{SD}(T) = \sqrt{\text{Var}(A)} = 14.4 \quad (3 \text{ s.f.})
\]

All of these follow from rescaling properties for random variables.